MEET SWISS INFOSEC!
SWISS INFOSEC AG

WENN ALGORITHMEN ZAUBERN: ECHTE SICHERHEIT UND DATENSCHUTZ FÜR KI

Thomas Nydegger, Managing Consultant

23. Juni 2025

ABLAUF

1. Risiken und Herausforderungen
2. Regulatorische Entwicklungen
3. Best Practices und Empfehlungen
4. Ausblick

RISIKEN UND HERAUS-FORDERUNGEN

Wie Elektrizität einst alle Lebensbereiche veränderte, transformiert KI nun ganze Branchen. Jede Firma sollte das Potential erkennen, um nicht den Anschluss zu verpassen.

"KI ist die neue Elektrizität…„

Andrew Ng, Stanford-Professor und KI-Pionier

SWISS INFOSEC

ECHTE SICHERHEIT UND DATENSCHUTZ FÜR KI
RISIKEN UND HERAUSFORDERUNGEN

INFORMATIONSSICHERHEIT

- Cyberrisiken
- Modellrisiken
- fehlende Verfügbarkeit
- Integritätsverlust
- Verletzung der Vertraulichkeit
- Abhängigkeit von Drittanbietern
- fehlende Governance
- Technologieüberalterung
- etc.

DATENSCHUTZ

- unrechtmässige Datenbear-beitung
- fehlende Transparenz / Aufklärung
- Identitätsmissbrauch
- unzureichende Wahrung der Rechte betroffener Personen
- automatisierte Einzelentscheidungen ohne menschliche Kontrolle
- Weitergabe ohne Kontrolle
- fehlende Anonymisierung
- etc.

VERANTWORTUNG

- Desinformation
- Diskriminierung
- Urheberrecht
- Produktehaftung
- unerlaubte Handlung
- Arbeitsverhältnis
- Vertrauenshaftung
- Reputationsrisiken
- Aufsichts- und Kontrollpflichten
- etc.

NEUE ANGRIFFSVEKTOREN ÜBER KI

- **Prompt Injection**: manipulative Eingaben zur Erzeugung ungewollter Anweisungen
- **Training Data Poisoning**: gezieltes Füttern von Trainingsdaten mit falschen oder manipulativen Informationen
- **Model Inversion**: Rekonstruktion von sensiblen Daten (z.B. Persönlichkeitsmerkmale) aus trainierten Modellen
- **Adversarial Examples**: Manipulation von Eingabedaten zwecks Täuschung von Menschen
- **Jailbreaking**: Aushebelung von Sicherheitsbeschränkungen durch verschachtelte oder kontextuelle Prompts
- **Model Hijacking**: gezielte Übernahme oder Manipulation von KI-Modellen durch Angreifer
- **Model-Theft / Model Extraction**: gezielte Abfrage und Nachbildung von KI-Modellen
- **Prompt Leaking**: gezielte Manipulation von KI-Modellen zur Offenlegung von internen Anweisungen, Systemprompts oder Verhaltensregeln
- **Synthetic Identity Fraud**: Erzeugung von gefälschten Identitäten durch Kombination von Deepfakes und KI-generierten Dokumenten

REGULATORISCHE ENTWICKLUNGEN

INTERNATIONALE ENTWICKLUNGEN

- **International**:
  - **OECD** Recommendation of the Council on Artificial Intelligence (2019/2024)
    - Förderung des menschlichen Wohlergehens und Förderung eines nachhaltigen Wachstums
    - Berücksichtigung der Menschenrechte, der Rechtsstaatlichkeit und der Demokratie
    - Informationen über das Funktionieren von KI-Systemen sollen zugänglich, erklärbar und überprüfbar sein.
    - KI-Systeme sollen technisch robust, manipulationssicher und resilient gegen Angriffe oder Ausfälle sein.
    - Entwickler, Anbieter und Nutzer von KI-Systemen sollen für deren Auswirkungen verantwortlich sein.
  - **ISO/IEC 42001:2023**: Artificial Intelligence Management System (2023)
  - **NIST**: AI Risk Management Framework (2023)

INTERNATIONALE ENTWICKLUNGEN

- **Europäische Union**: Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (2024)
- **Europarat**: Framework Convention on Artificial Itelligence and Human Rights, Democracy and the Rule of Law (2024)
- **USA**: Executive Order on Safe, Secure, and Trustworthy AI (2023) (am 20.01.2025 wieder aufgehoben)
- **China**: Ethical Guidelines for New Generation AI (2021), Interim Measures for the Management of Generative AI Services (2023), Deep Synthesis Regulation (2023) und Measures for Labeling of AI-Generated Synthetic Content (2025)
- **UK**: White Paper „AI Regulation: A Pro-Innovation Approach“ (2023)
- **Australien**: AI Ethics Framework (2019)
- **Japan**: Governance Guidelines for Deployment of AI (2022)

SCHWEIZ – KI-REGULIERUNG (GEPLANT BIS 2026)

Historischer Kontext:

- Bis 2022: Schweiz verfolgt zurückhaltenden Ansatz, bestehende Gesetze ausreichend.
- Februar 2023: KI als Schwerpunkt des Bundesrates festgelegt.
- Februar 2025: Entscheid zur Ratifizierung der Europarat KI-Konvention.

Ziele & Einfluss:

- Anpassung nationaler Gesetze zur Einhaltung internationaler Standards.
- Sicherstellung Wettbewerbsfähigkeit und Marktzugang zur EU (AI Act).

Aktueller Status:

- Erarbeitung Gesetzesentwurf bis Ende 2026.
- Orientierung an internationalen Normen (Europarat).

Rolle	Anforderungen
Geplant (Anlehnung an Europarat)	• Transparenzpflichten, klare Verantwortlichkeiten • Risikomanagement, Bias-Vermeidung, Haftungsregeln • Rechtsschutz bei automatisierten Entscheidungen
Betroffen sind	• Entwickler (Dokumentation, Tests, Zertifizierungen) • Betreiber (Risikomanagement, Compliance-Pflichten, Governance) • Benutzer (Rechte bei KI-Entscheidungen)

KI-REGULIERUNG, AKTUELLE AKTIVITÄTEN IN DER SCHWEIZ

Datenschutz

Das totalrevidierte Schweizer Datenschutzgesetz (DSG) gilt auch für KI-Anwendungen, insbesondere wenn Personendaten automatisiert verarbeitet werden. KI-Systeme müssen datenschutzkonform gestaltet und betrieben werden.

Urheberrecht

Derzeit werden im Bereich des Urheberrechts (URG) entscheidende Fragen zum Thema „Training mit urheberrechtlich geschützten Werken“ und zum Thema „urheberrechtlicher Schutz von KI-Ergebnissen“ gestellt. Von Bedeutung ist auch die Frage, welcher individueller Charakter einem Prompt zukommt, um als neue geistige Schöpfung zu gelten.

Produktehaftpflichtrecht

Bisher wurde Software nicht unter die Herstellerhaftung nach dem Produktehaftpflichtgesetz (PrHG) unterstellt. Vor dem Hintergrund der Revision der EU-Richtlinie über die Haftung für fehlerhafte Produkte wird aber ein Modernisierungsbedarf des PrHG erkannt.

Verkehr (autonome Fahrzeuge)

Der Einsatz autonomer Fahrzeuge ist rechtlich geregelt, etwa durch das Strassenverkehrsgesetz und spezifische Zulassungsanforderungen. Die Schweiz hat 2025 grünes Licht für den Einsatz autonomer Fahrzeuge auf öffentlichen Strassen gegeben.

Digitale Plattformen und Medien

Vorgaben gegen Desinformation, Deepfakes und gewalttätige Inhalte auf sozialen Medien werden aktuell vorbereitet. Ein neues Gesetz für digitale Plattformen ist in Planung, das auch KI-gestützte Inhalte adressiert. Der BR hat aber das Geschäft Bundesgesetz über Kommunikationsplattformen und Suchmaschinen (KomPG) im April 2025 bis auf weiteres verschoben.

Arbeitsschutz und Diskriminierung

Arbeitsrechtliche Vorschriften und das Gleichstellungsgesetz gelten auch für KI-gestützte Personalentscheidungen, um Diskriminierung zu verhindern. Der Schutz vor Diskriminierung soll erweitert werden.

BEST PRACTICES UND EMPFEHLUNGEN

TECHNISCHE RISIKEN VON AI

Risiko	Beispiel	Lösung
Adversarial Attacks (Manipulation)	Geänderte Bilder führen zu falschen Diagnosen in medizinischen KI-Modellen.	Robustheitstests, Anomalieerkennung, sichere Modellentwicklung.
Modell-Drift (Veränderungen über Zeit)	Veraltete Daten führen dazu, dass AI neue Krankheitsbilder nicht erkennt.	Regelmässige Modellüberprüfung, kontinuierliches Training.
Halluzinationen & Fehlinformationen	ChatGPT gibt falsche medizinische Ratschläge zu Medikamentenwechselwirkungen.	Strenge Validierung, Fact-Checking, AI-Überwachung durch Experten.

DATENSCHUTZRISIKEN IN AI-SYSTEMEN

Risiko	Beispiel	Lösung
Bias & Diskriminierung	AI-Bewerbungssystem bevorzugt Männer aufgrund unausgewogener Trainingsdaten.	Diverse Trainingsdaten, Bias-Checks, Fairness-Algorithmen.
Datenlecks & Datenschutzverletzungen	AI-Chatbot speichert sensible Patientendaten unverschlüsselt.	Datenmaskierung, Verschlüsselung, Zugriffskontrollen.
Mangelnde Datenintegrität & Vertrauensprobleme	Manipulierte medizinische Daten verursachen Fehldiagnosen.	Strenge Datenvalidierung, Audit-Trails, Protokollierung von Modell-Inputs.

ORGANISATORISCHE RISIKEN DURCH AI

Risiko	Beispiel	Lösung
Fehlende Verantwortlichkeiten	Autonomes Fahrzeug verursacht Unfall – Wer ist verantwortlich?	Klare AI-Governance, Verantwortlichkeitsregelungen, Dokumentation.
Übermässige Abhängigkeit von AI	Arzt verlässt sich ausschliesslich auf AI-Diagnosen, ignoriert eigene Erfahrung.	„Human-in-the-loop“-Ansätze, menschliche Validierung.
Unklare Compliance-Vorgaben & fehlende Regulierung	Unternehmen entwickelt KI ohne Berücksichtigung des EU AI Act.	Einhaltung regulatorischer Anforderungen (ISO 42001, EU AI Act, DSG).

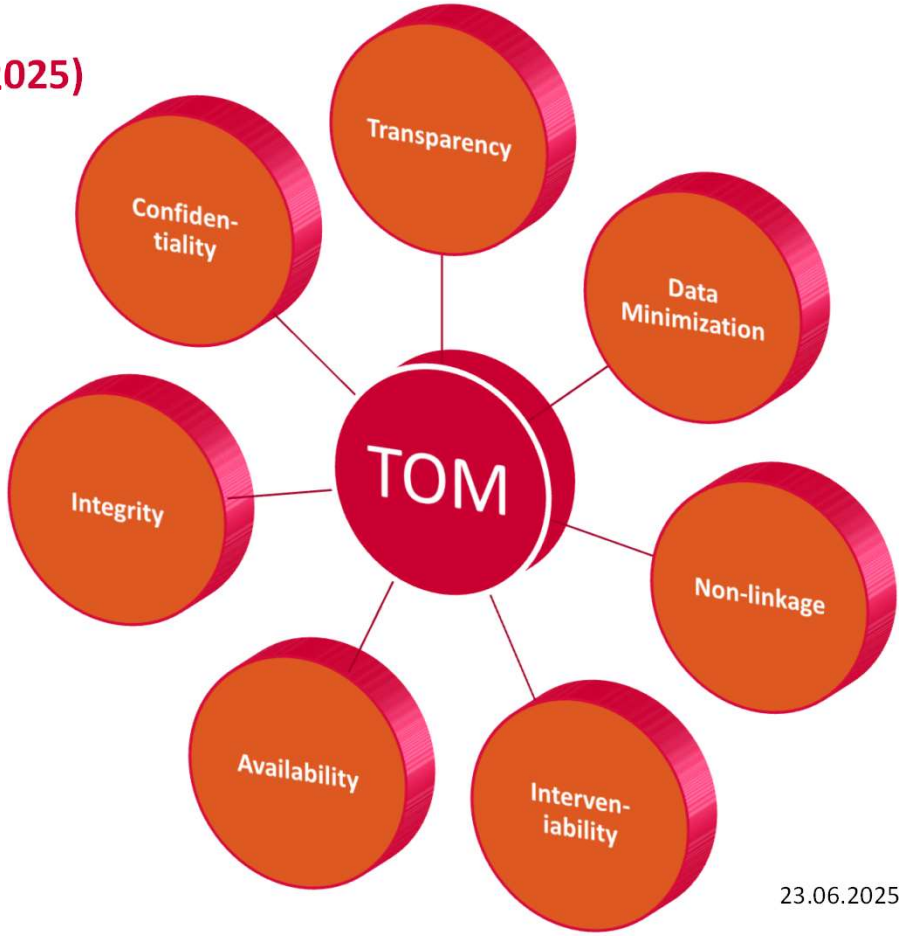
DSK-ORIENTIERUNGSHILFE FÜR KI-ENTWICKLER (JUNI 2025)

1. Design
(Auswahl der Daten, Datensammlung)

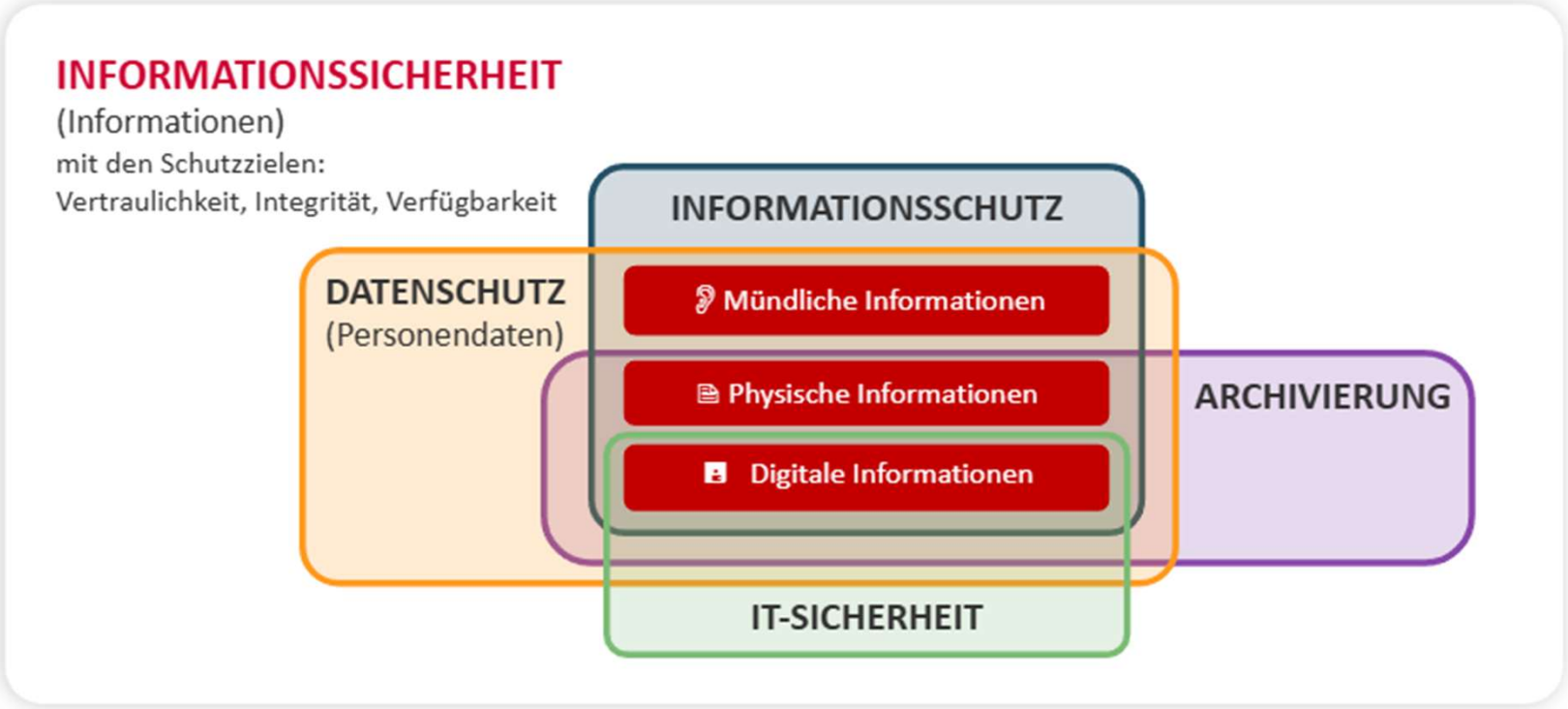
2. Entwicklung
(Datenaufbereitung, Training und Validierung)

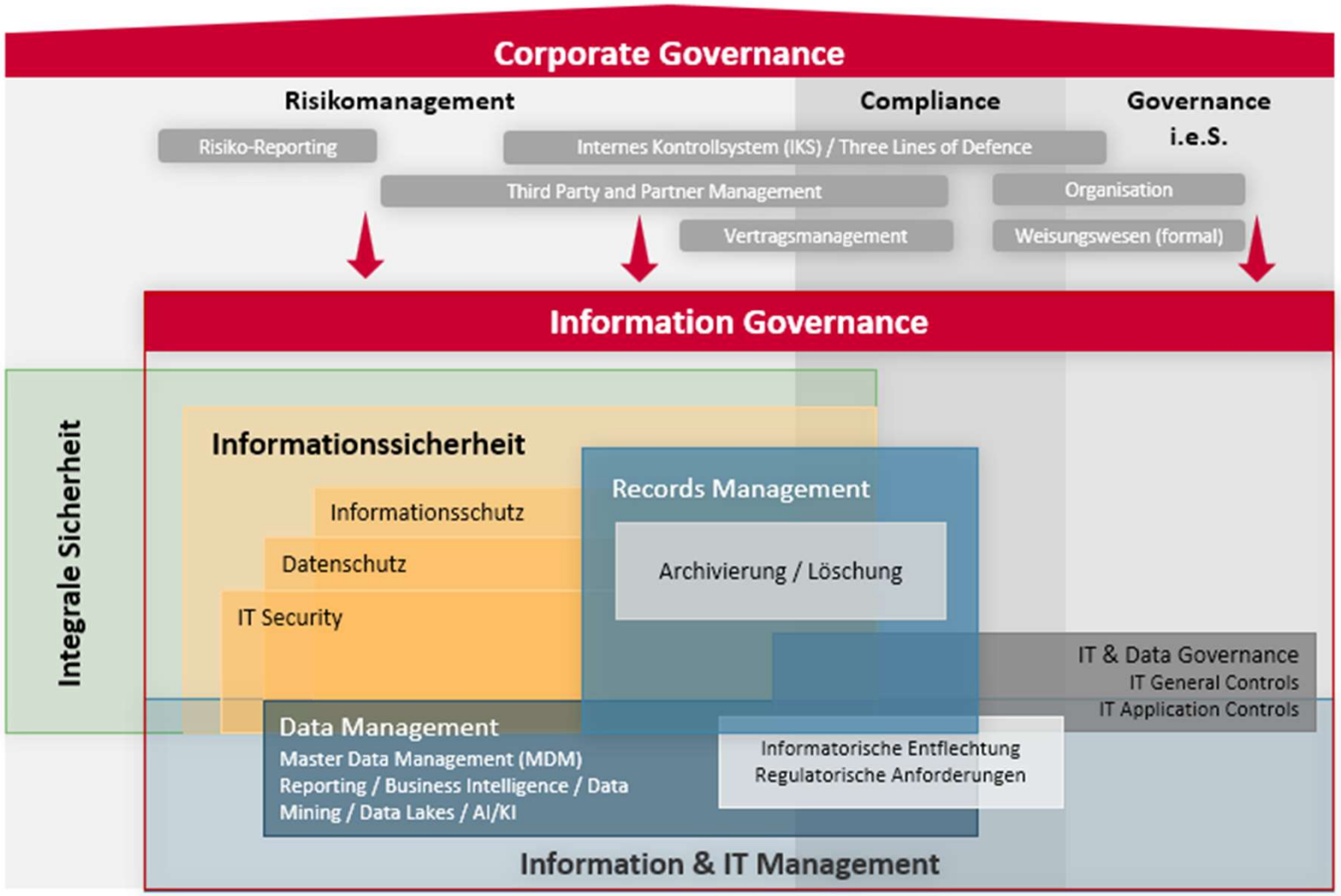
3. Einführung
(Softwareverteilung inkl. Updates)

4. **Betrieb und Monitoring**



KI UND INFORMATIONSSICHERHEIT




Corporate Governance
Risikomanagement
Compliance
Governance i.e.S.
Risiko-Reporting
Internes Kontrollsystem (IKS) / Three Lines of Defence
Third Party and Partner Management
Organisation
Vertragsmanagement
Weisungswesen (formal)
Information Governance
Integrale Sicherheit
Informationssicherheit
Informationsschutz
Datenschutz
IT Security
Records Management
Archivierung / Löschung
IT & Data Governance
IT General Controls
IT Application Controls
Data Management
Master Data Management (MDM)
Reporting / Business Intelligence / Data Mining / Data Lakes / AI/KI
Informatorische Entflechtung
Regulatorische Anforderungen
Information & IT Management

VERTRAUENSWÜRDIGE KI

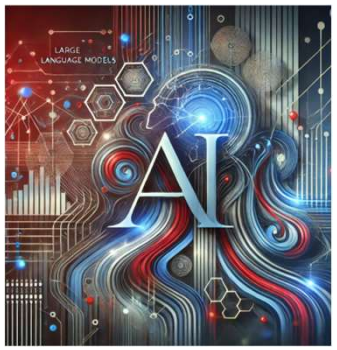
1. Transparenz
- Erklärbarkeit und Nachvollziehbarkeit
- Zugänglichkeit von Informationen
- Auditierbarkeit

2. Fairness und Nichtdiskriminierung
- Vermeidung von Bias
- Fairness in Entscheidungen
- Gerechte Nutzungsverteilung

3. Datenschutz und Datensicherheit
- Einhaltung der Datenschutzgesetze
- Datenminimierung
- Sichere Speicherung und Verarbeitung

4. Verantwortlichkeit
- Klar definierte Zuständigkeiten
- Nachvollziehbarkeit der Ergebnisse
- Regelungen für Notfallsituationen

5. Robustheit und Sicherheit
- Robustheit gegen Fehler und Angriffe
- Resilienz und Redundanz
- Regelmässige Überprüfung und Aktualisierung



6. Ethische und gesellschaftliche Auswirkungen
- Vermeidung von Schaden
- Förderung des Gemeinwohls
- Bewusstsein für gesellschaftliche Folgen

7. Human-in-the-Loop (HITL)
- Menschliche Kontrolle und Überwachung
- Mensch aktiv bei Entscheidungen eingebunden

8. Ethische Richtlinien und Gesetzgebung
- Einhaltung internationaler Standards
- Verankerung in Gesetzen und Richtlinien

Siehe bspw. AI Cloud Service Compliance Criteria Catalog (AIC4), 2024

DIE 7 GOLDENEN REGELN FÜR DEN VERANTWORTUNGSVOLLEN EINSATZ VON KÜNSTLICHER INTELLIGENZ (KI)

1. **Verantwortung regeln und übernehmen – Gesetze einhalten**
   Sei dir der Auswirkungen von KI-Systemen bewusst und stelle sicher, dass menschliche Kontrolle jederzeit gewährleistet ist.
2. **Transparenz herstellen**
   Sorge dafür, dass der Einsatz der KI nachvollziehbar ist. Dokumentiere den Einsatz von KI.
3. **Datenschutz und Privatsphäre wahren**
   Schütze die Daten und die Privatsphäre der Betroffenen. Halte strikte Datenschutz- und Sicherheitsprotokolle ein.
4. **Fairness und Gerechtigkeit gewährleisten**
   Entwickle und nutze KI so, dass sie frei von Diskriminierung und Vorurteilen ist und alle Menschen gleich behandelt werden.
5. **Sicherheit und Zuverlässigkeit garantieren**
   Sorge dafür, dass KI-Systeme (auch in komplexen Situationen) sicher und zuverlässig arbeiten und Risiken minimiert werden.
6. **Schaden und Missbrauch verhindern**
   Implementiere Schutzmechanismen, um sicherzustellen, dass KI-Systeme nicht schaden und nicht für schädliche oder illegale Zwecke missbraucht werden.
7. **Geistiges Eigentum und Rechte Dritter respektieren**
   Respektiere den Schutz des geistigen Eigentums und verletze nicht die Rechte Dritter.

CHECKLISTE UMSETZUNG

KI Vorgaben und Verantwortlichkeiten
- Benennung eines KI-Verantwortlichen/einer Fachgruppe
- Weisung KI Governance erstellen
  (gibt klare Vorgaben für den gesamten KI Lifecycle)
- Breite Abstimmung

Inventar
- Identifikation der existierenden KI-Systeme
- Prüfung der vertraglichen Grundlagen

Risikobewertung der KI-Systeme
- Klassifizierung der KI-Systeme
- Durchführung von Risikoanalysen / Datenschutzfolgeabschätzungen
- Laufende Nachführung Risikobehandlungsplan

Spezifisch KI-technische Anforderungen
- Datenqualität und -governance sicherstellen
- Transparenz und Erklärbarkeit implementieren
- Genauigkeit und Robustheit prüfen und nachweisen

Schulungen und Awareness

Dokumentation und Protokollierung
- Technische Dokumentationen/ Bearbeitungsreglemente
- Protokollierung sicherstellen
- Bereitstellung von Nachweisen

Governance und Compliance
- Interne Kontrollen implementieren
- Verfahren für neue KI-Funktionen und KI-Changes durchsetzen

Monitoring
- Laufende Überwachung der KI-Systeme

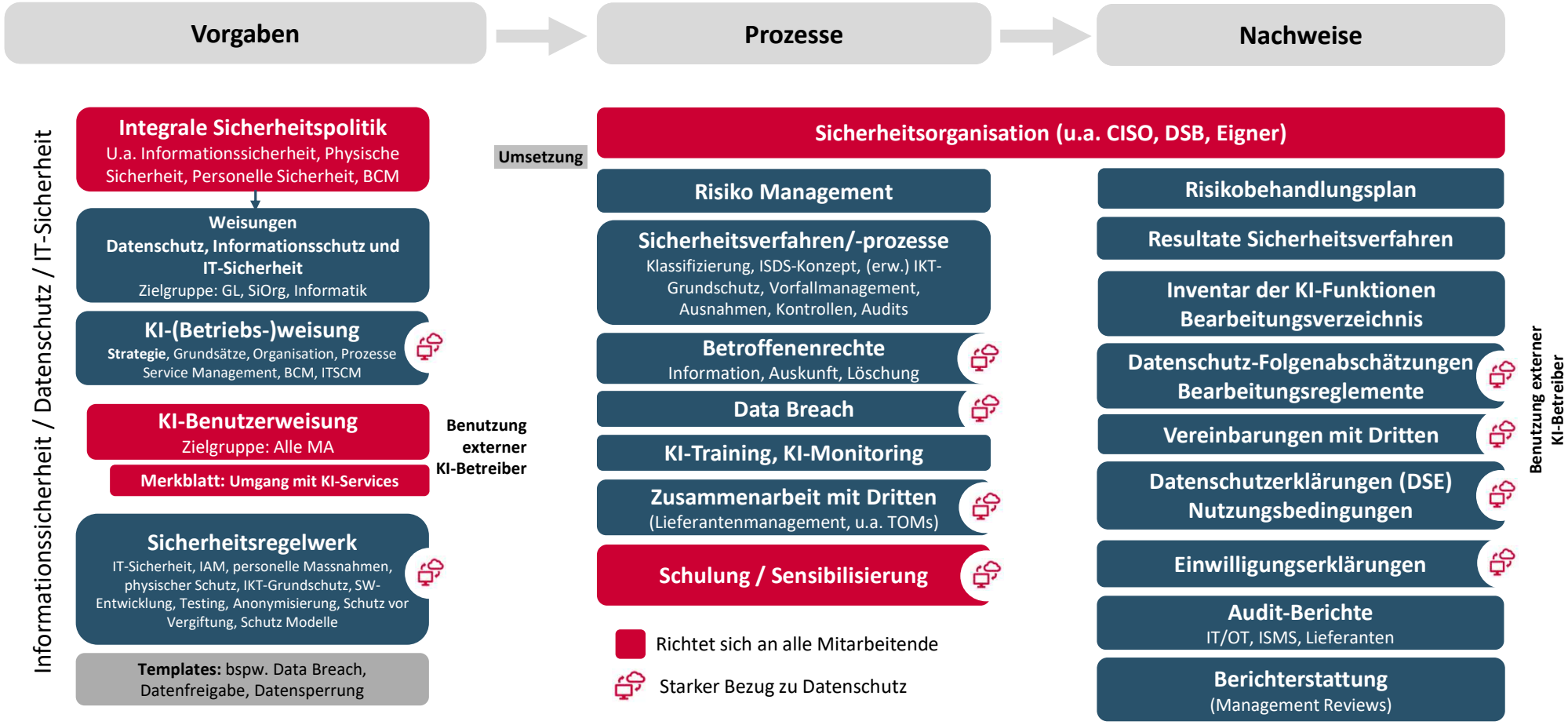
Betroffenenrechte / Transparenz
- Information an Nutzer sicherstellen
- Prozesse Betroffenenrechte implementieren

Durchführung von Audits

Notfallmassnahmen

SWISS INFOSEC

ECHTE SICHERHEIT UND DATENSCHUTZ FÜR KI
KI-HAUS, UMSETZUNGSEMPFEHLUNGEN



SWISS INFOSEC

ECHTE SICHERHEIT UND DATENSCHUTZ FÜR KI
FRAGEN

SWISS INFOSEC

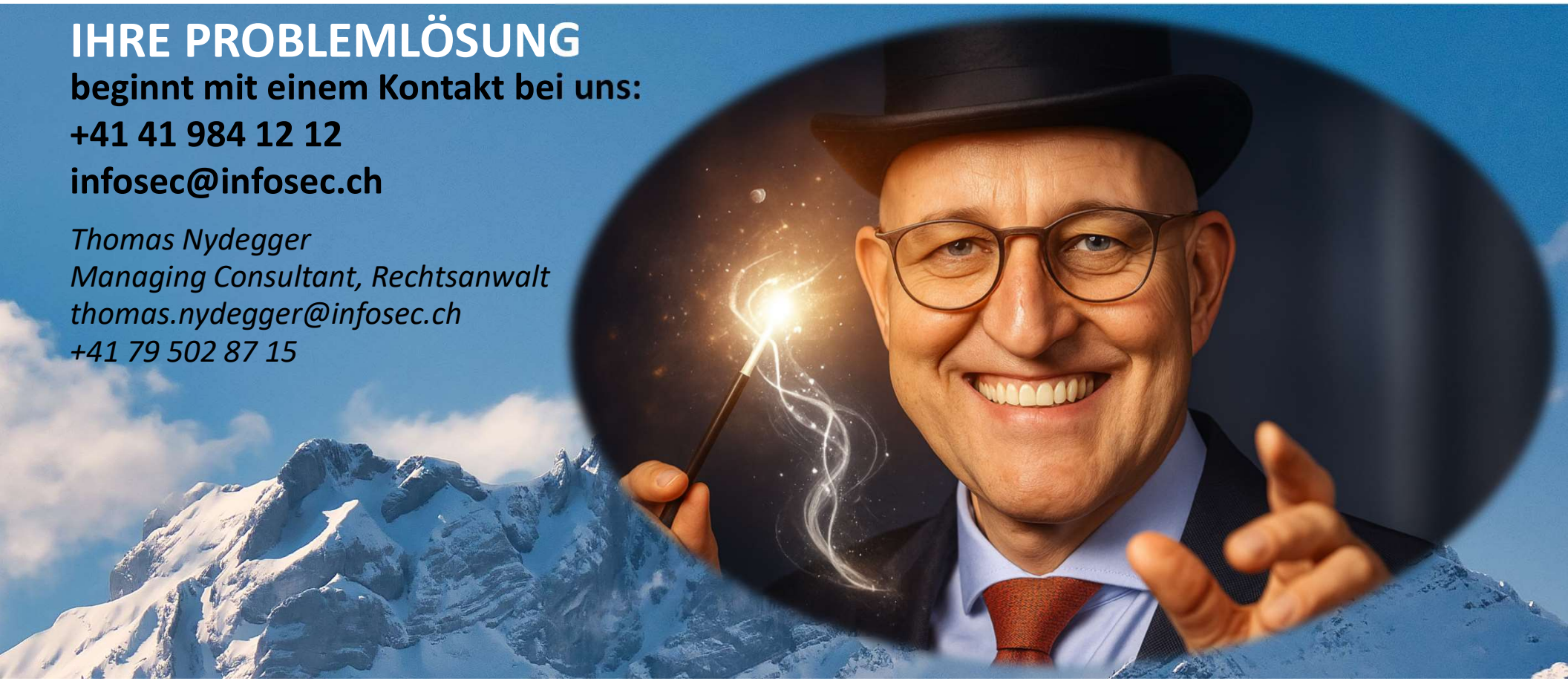
IHR KONTAKT
SWISS INFOSEC AG

IHRE PROBLEMLÖSUNG

beginnt mit einem Kontakt bei uns:

+41 41 984 12 12
infosec@infosec.ch

Thomas Nydegger
Managing Consultant, Rechtsanwalt
thomas.nydegger@infosec.ch
+41 79 502 87 15



SWISS INFOSEC

VIELEN
DANK

MELDEN SIE SICH JETZT AN FÜR DIE KOSTENLOSE FACHVERANSTALTUNG

MEET SWISS INFOSEC!
Sicherheit im Fokus

Zürich Flughafen
13 bis 17 Uhr, anschliessend Apéro
www.infosec.ch/msi

Haben Sie schon den kostenlosen Newsletter abonniert?
www.infosec.ch/news